AHRQ Comparative Effectiveness Review Surveillance Program

CER #59:
Screening for Glaucoma: Comparative Effectiveness

Original release date:
April, 2012

Surveillance Report:
May, 2013

Key Findings:

- A number of new studies were identified assessing the sensitivity and specificity of the Heidelberg Retina Tomograph 3, optical coherence tomography (OCT), and scanning laser polarimetry. All of the original conclusions regarding screening tests remain valid with the possible exception of the conclusion regarding OCT.
- Apart from the individual tests addressed by the original key questions, several studies and expert comments addressed the use of a combination of tests for screening purposes.

Summary Decision

This CER's priority for updating is **Low**

Authors:
Sydne Newberry, PhD
Aneesa Motala, BA
Susanne Hempel, PhD
Jennifer Schneider Chafen, MS, MD
Margaret Maglione, MPP
Roberta Shanman, MS
Paul Shekelle, MD, PhD

None of the investigators has any affiliations or financial involvement that conflicts with the material presented in this report.

Acknowledgments

The authors gratefully acknowledge the following individuals for their contributions to this project:

Subject Matter Experts

Anne Coleman, MD
Jules Stein Eye Institute, University of California-Los Angeles
Los Angeles, CA

Ann-Margaret Ervin, PhD
Johns Hopkins Bloomberg School of Public Health
Baltimore, MD

JoAnn Giaconni
Jules Stein Eye Institute, University of California- Los Angeles
Los Angeles, CA

Murray Fingeret, O.D., FAAO
Department of Veterans Administration New York Harbor Health Care System
State University of New York College of Optometry
New York, NY

Flora Lum, MD
American Academy of Ophthalmology
San Francisco, CA

Timothy Wilt, MD
Veterans Affairs, Minneapolis
Minnesota Evidence-based Practice Center
Minneapolis, MN

Linda Zangwill, PhD
Shiley Eye Institute, University of California- San Diego
La Jolla, CA

Contents

Screening for Glaucoma: Comparative Effectiveness

1. Introduction

Comparative Effectiveness Review (CER) #59, Screening for Glaucoma: Comparative Effectiveness, was released in April 2012.[1] It was therefore due for a surveillance assessment in October, 2012. At that time, we contacted experts involved in the original CER and subject experts to get their opinions as to whether the conclusions had changed and need to be updated. We also conducted an update electronic literature search. Every month since the CER's original release, we received any FDA updates on the included treatments and tests.

2. Methods

2.1 Literature Searches

Using the search strategy employed for the original report, we conducted a limited literature search of Medline for the years August 2011-December 26, 2012. This search included five high-profile general medical interest journals (Annals of Internal Medicine, British Medical Journal, Journal of the American Medical Association, Lancet, and the New England Journal of Medicine) and six specialty journals (Journal of Glaucoma, American Journal of Ophthalmology, Ophthalmology, Investigative Ophthalmology & Visual Science, Archives of Ophthalmology, and the British Journal of Ophthalmology). The specialty journals were the most highly represented among the references for the original report. Appendix A includes the search methodology for this topic.

2.2 Study selection

In general we used the same inclusion and exclusion criteria as the original CER. In particular, study populations were required to include glaucoma suspects (those at high risk for glaucoma) and could not be limited to healthy participants and those with diagnosed glaucoma.

2.3 Expert Opinion

We shared the conclusions of the original report with 10 experts in the field (including the original project leader, all original technical expert panel (TEP) members, peer reviewers, and local content experts for their assessment of the need to update the report and their recommendations of any relevant new studies; 7 subject matter experts responded, including the project lead. Appendix C shows the questionnaire matrix that was sent to the experts.

2.4 Check for qualitative and quantitative signals

After abstracting the study conditions and findings for each new included study into an evidence table, we assessed whether the new findings provided a signal according to the Ottawa Method and/or the RAND Method, suggesting the need for an update. The criteria are listed in the table below.[2, 3]

	Ottawa Method
	Ottawa Qualitative Criteria for Signals of Potentially Invalidating Changes in Evidence
A1	Opposing findings: A pivotal trial or systematic review (or guidelines) including at least one new trial that characterized the treatment in terms opposite to those used earlier.
A2	Substantial harm: A pivotal trial or systematic review (or guidelines) whose results called into question the use of the treatment based on evidence of harm or that did not proscribe use entirely but did potentially affect clinical decision making.
A3	A superior new treatment: A pivotal trial or systematic review (or guidelines) whose results identified another treatment as significantly superior to the one evaluated in the original review, based on efficacy or harm.
	Criteria for Signals of Major Changes in Evidence
A4	Important changes in effectiveness short of "opposing findings"
A5	Clinically important expansion of treatment
A6	Clinically important caveat
A7	Opposing findings from discordant meta-analysis or nonpivotal trial
	Quantitative Criteria for Signals of Potentially Invalidating Changes in Evidence
B1	A change in statistical significance (from nonsignificant to significant)
B2	A change in relative effect size of at least 50 percent
	RAND Method Indications for the Need for an Update
1	Original conclusion is still valid and this portion of the original report does not need updating
2	Original conclusion is possibly out of date and this portion of the original report may need updating
3	Original conclusion is probably out of date and this portion of the original report may need updating
4	Original conclusion is out of date

2.5 Compilation of Findings and Conclusions

For this assessment we constructed a summary table that included the key questions, the original conclusions, and the findings of the new literature search, the expert assessments, and any FDA reports that pertained to each key question. To assess the conclusions in terms of the evidence that they might need updating, we used the 4-category scheme described in the table above for the RAND Method.

In making the decision to classify a CER conclusion into one category or another, we used the following factors when making our assessments:

- If we found no new evidence or only confirmatory evidence and all responding experts assessed the CER conclusion as still valid, we classified the CER conclusion as still valid.
- If we found some new evidence that might change the CER conclusion, and /or a minority of responding experts assessed the CER conclusion as having new evidence that

might change the conclusion, then we classified the CER conclusion as possibly out of date.
- If we found substantial new evidence that might change the CER conclusion, and/or a majority of responding experts assessed the CER conclusion as having new evidence that might change the conclusion, then we classified the CER conclusion as probably out of date.
- If we found new evidence that rendered the CER conclusion out of date or no longer applicable, we classified the CER conclusion as out of date. Recognizing that our literature searches were limited, we reserved this category only for situations where a limited search would produce prima facie evidence that a conclusion was out of date, such as the withdrawal of a drug or surgical device from the market, a black box warning from FDA, etc.

2.6 Determining Priority for Updating

We used the following two criteria in making our final conclusion for this CER:

- How much of the CER is possibly, probably, or certainly out of date?
- How out of date is that portion of the CER? For example, would the potential changes to the conclusions involve refinement of original estimates or do the potential changes mean some therapies are no longer favored or may not exist? Is the portion of the CER that is probably or certainly out of date an issue of safety (a drug withdrawn from the market, a black box warning) or the availability of a new drug within class (the latter being less of a signal to update than the former)?

3. Results

3.1 Search

The literature search identified 248 titles. After title and abstract review, we further reviewed the full text of 30 journal articles. The remaining titles were rejected because they clearly did not meet inclusion criteria for any of the review questions. In addition to the electronic database searches, we followed up suggestions from the topic experts for studies not already included in the original report. We reference-mined articles that met inclusion criteria as well as systematic reviews identified by the literature searches to identify additional articles that may have been published since the publication of the report.

Thus, 56 articles went on to full text review. Of these, 37 articles were rejected because they did not meet the inclusion criteria of the original report. The remaining 19 articles were abstracted into an evidence table (Appendix B) for this assessment.[4-22]

3.2 Expert Opinion

Of the seven experts who provided feedback on the original conclusions by completed the questionnaire matrix, most agreed that most of the conclusions remained up to date, with two exceptions. Three reviewers provided a number of new references regarding the use of newer

technologies, particularly spectral domain (SD) optical coherence tomography (OCT) or time domain (TD) OCT for glaucoma diagnosis. A portion of these references considered the potential use of OCT for glaucoma screening, whereas the remainder considered only the sensitivity and specificity for diagnosis among various subpopulations, including only healthy individuals and individuals with confirmed glaucoma diagnoses among the study participants. The latter studies were excluded. A fourth reviewer noted the proliferation of studies of OCT but questioned its utility for screening settings. Thus, for the question of the predictive value of this modality, we noted that the "Original conclusion is possibly out of date and this portion of the original report may need updating."

Two of the experts also commented, apart from their responses regarding conclusions to individual questions, that future screening programs would likely use a combination of two or more diagnostic modalities, and one of these reviewers provided several references. The latter included the results of a Delphi process on designing the components of a screening program. However, to date, the literature on multistep screening for glaucoma does not appear to merit a new review.

3.3 Identifying qualitative and quantitative signals

Table 1 shows the original key questions, the conclusions of the original report, the results of the literature and drug database searches, the experts' assessments, the recommendations of the Southern California Evidence-based Practice Center (SCEPC) regarding the need for update, and qualitative signals.

Table 1: Summary Table

Conclusions From CER Executive Summary	RAND Literature Search	FDA/Health Canada/MHRA (UK)	Expert Opinion EPC Investigator Other Experts	Conclusion from SCEPC
KQ1a: Does a screening-based program for open-angle glaucoma lead to less visual impairment when compared with no screening program? KQ1b: How does visual impairment vary when comparing different screening-based programs for open-angle glaucoma?				
We did not identify any study that addressed whether participation in an OAG screening-based program leads to less visual impairment when compared with no screening or another screening-based program.	No new studies were identified.	None found	6 reviewers stated the conclusion is still supported by the evidence. 1 of these 6 reviewers cited a potentially relevant study. 1 reviewer said s/he did not know.	Original conclusion is still valid and this portion of the original report does not need updating.
KQ2a: Does a screening-based program for open-angle glaucoma lead to improvements in patient-reported outcomes when compared to no screening? KQ2b: How do patient-reported outcomes vary when comparing different screening-based programs for open angle glaucoma?				
We did not identify any study that addressed whether participation in an OAG screening-based program leads to improvements in patient-reported outcomes when compared with no screening or another screening-based program.	No new studies were identified.	None found	5 reviewers stated the conclusion is still supported by the evidence. 2 reviewers said they did not know.	Original conclusion is still valid and this portion of the original report does not need updating.
KQ 3: What is the predictive value of screening tests for open angle glaucoma?				
Evidence for this question came from two sources: a 2007 systematic review of diagnostic test accuracy for OAG (with subsequent diagnosis on followup as the reference standard) and original studies published subsequent to that review. No RCTs	One study suggested by an expert[19] on the use of the swinging flashlight technique to assess relative afferent pupillary defect (RAPD) and another study about RAPD[5] found high specificity but poor (0%) to moderate (66.7%) sensitivity for screening. Exclusion of prior cataract	None found	4 reviewers stated the conclusion is still supported by the evidence. 1 reviewer cited potentially relevant studies	Original conclusion is still valid and this portion of the original report does not need updating.

Conclusions From CER Executive Summary	RAND Literature Search	FDA/Health Canada/MHRA (UK)	Expert Opinion EPC Investigator Other Experts	Conclusion from SCEPC
specificity are reported by test, based on the evidence review and the subsequent literature. The authors of the original SR noted significant heterogeneity and high risk of bias for most groups of studies.	specificity as well as PPV. (other screening tests are described below)		did not know.	
Tests of Optic Nerve Structure **Heidelberg Retina Tomograph II** **Evidence From Burr et al., 2007**. pooled sensitivity: 86 percent (95% credible interval [CrI], 55 to 97) pooled specificity 89 percent (95% CrI, 66 to 98). **Evidence From Primary Studies.** Of 17 primary studies, 2 specifically focused on detecting early or moderate glaucoma. Sensitivity across 12 parameters: 47.1 percent (RNFL cross-sectional area) to 74.3 percent (linear cup/disc area ratio), Specificity: 47.1 percent (mean RNFL thickness) to 71.4 percent (cup shape measure). 15 studies compared HRT II with other devices. HRT II was found not to perform as well as GDx VCC, OCT, or FDT. HRT II and HRT III were found to have a similar diagnostic profile. Three of the included studies concluded that HRT II was not an appropriate tool	No new studies were identified.	None found	5 reviewers stated the conclusion is still supported by the evidence, although 1 stated that HRT II and III are now combined. 1 reviewer cited potentially relevant studies. 1 reviewer said s/he did not know.	Original conclusion is still valid and this portion of the original report does not need updating.
Heidelberg Retina Tomograph III **Evidence from Burr et al.:** not included **Evidence From Primary Studies.** Eleven studies examined the	One new study used Moorfield's Regression Analysis (MRA) with HRT.[21] The specificity precludes its use as a screening tool but may be useful in older individuals for whom high quality scans	None found	5 reviewers stated the conclusion is still supported by the evidence. 1 reviewer said s/he did	Original conclusion is still valid and this portion of the original report does not need updating.

Conclusions From CER Executive Summary	RAND Literature Search	FDA/Health Canada/MHRA (UK)	Expert Opinion EPC Investigator Other Experts	Conclusion from SCEPC
diagnostic accuracy of HRT III. The sensitivity of the Glaucoma Probability Score (HRT III) for distinguishing eyes with early field loss from healthy eyes was 67.9 percent, and that of the Moorfields Regression Analysis was 71.9 (at a fixed specificity of 92 percent). The investigators concluded that "the relative ease and sensitivity of the operator-independent Glaucoma Probability Score function of the HRT III may facilitate glaucoma screening." In another study, sensitivity (parameter: reference height) ranged from 4 to 70 percent when holding the specificity of the test constant at 95 percent.	are available. Another study in a primarily Caribbean and African population in Canada found that glaucoma probability score (GPS) with HRTIII is less specific but more sensitive than MRA with HRTIII and as a screening tool has the advantage of being independent of contour lines.[9] Another study by the same group found that in high-risk populations, rim area to disc area asymmetry ratio (RADAAR) had a higher specificity than MRA but that for borderline cases on MRA, combining the 2 increased specificity.[10] Another study that was attempting to establish normative values for HRT3 variables and to develop HRT-based criteria for glaucoma for population-based research found that at a high specificity, sensitivity is low, although not as low as vertical CDR. Adjusting HRT3 for linear cup to disc ratio (LCDR) seems to be the most suitable variable to develop criteria for glaucomatous optic neuropathy for epidemiological purposes.[13]		not know. 1reviewer did not respond.	
<u>Ophthalmoscopy</u> **Evidence From Burr et al., 2007**. Seven studies addressed the diagnostic accuracy of ophthalmoscopy. Pooled sensitivity for five studies with common cutoff point of a vertical cup-to-disc ratio greater than or equal to 0.7: 60 percent (95% CrI, 34 to 82) Specificity: 94 percent (95% CrI, 76 to	No new studies were identified.	None found	6 reviewers stated the conclusion is still supported by the evidence. 1 reviewer said s/he did not know.	Original conclusion is still valid and this portion of the original report does not need updating.

Conclusions From CER Executive Summary	RAND Literature Search	FDA/Health Canada/MHRA (UK)	Expert Opinion EPC Investigator Other Experts	Conclusion from SCEPC
99). The diagnostic odds ratio (DOR) was 25.7 (95% CrI, 5.79 to 109.50), suggesting a 26-fold higher odds of a positive test among those with glaucoma than those without glaucoma.				
Optical Coherence Tomography (OCT) **Evidence From Primary Studies**. 47 included studies that investigated the diagnostic accuracy of OCT: 34 considered the Stratus OCT, 10 included the Cirrus OCT, 6 considered the RTVue OCT, 2 included the Spectralis OCT, 2 examined the OTI OCT, and 1 included the OTI Spectral OCT/SLO. Stratus OCT studies at high risk of spectrum bias because those with known disease as well as those with healthy eyes were enrolled in the studies. Range of specificity: 66 to 100 percent.	Numerous studies on OCT were identified. One study assessed the effect of optic disc size on the diagnostic accuracy of RTVue spectral domain (SD) OCT and found that optic disk size did not influence the AUC of any of the SD-OCT scanning protocols.[18] Another study by the same group[7] found that the effectiveness of most SD-OCT parameters to detect glaucoma significantly decreased when a clinically relevant group of glaucoma suspects was included, rather than relying on a population with known glaucoma and a healthy population. Another study found that the macular hemifield test showed better sensitivity than RNFL thickness measures in early glaucoma patients, with the same specificity.[8] Another study found that RNFL assessment with SD-OCT performed well in detecting preperimetric glaucomatous damage in glaucoma suspects and performed better than CSLO, as measured by AUC, however, the authors conclude that the test may not be the most appropriate for screening for	None found	3 reviewers stated the conclusion is still supported by the evidence, although 1 reviewer questioned the consideration of SD-OCT for screening. 3 reviewers cited potentially relevant studies. 1 reviewer said s/he did not know.	Original conclusion is possibly out of date and this portion of the original report may need updating.

Conclusions From CER Executive Summary	RAND Literature Search	FDA/Health Canada/MHRA (UK)	Expert Opinion EPC Investigator Other Experts	Conclusion from SCEPC
	glaucomatous visual field loss in the general population.[22] A study of Time domain (TD)-OCT[15] found that RNFL parameters discriminate ocular hypertension patients and those with glaucoma in the other eye better than did optic nerve head (ONH) parameters. ONH was better at discriminating preperimetric from normal eyes. A study that compared TD-OCT and SD-OCT found that AUC were comparable for both the SD and TD-OCT. RFNL thickness measures correlate well but do not have clinically acceptable agreement and therefore are not interchangeable.[11] A Taiwanese study showed that both SD- and TD-OCT had comparable diagnostic power, e.g., to distinguish early glaucoma, OH, glaucoma suspects, POAG, and PACG in a Taiwanese population.[4] Another study that compared SD-OCT and TD-OCT found that TD-OCT had better specificity and PPV. SD-OCT had only 28% sensitivity among early patients: therefore TD-OCT better for screening.[20]			
Optic Disc Photography **Evidence From Burr et al., 2007.** (6 studies) Sensitivity: 65 to 77 percent, Specificity: 59 to 98 percent. **Evidence From Primary Studies.** two studies of the diagnostic accuracy	No new studies were identified.	None found	6 reviewers stated the conclusion is still supported by the evidence. 1 reviewer said s/he did not know.	Original conclusion is still valid and this portion of the original report does not need updating.

Conclusions From CER Executive Summary	RAND Literature Search	FDA/Health Canada/MHRA (UK)	Expert Opinion EPC Investigator Other Experts	Conclusion from SCEPC
of optic disc photography and one study of cup-to-disc ratio measurement as measured by an ophthalmologist using a slit-lamp biomicroscope and 78 Diopter lens. Danesh- Meyer et al. (2006) included participants with OAG as well as glaucoma suspects and healthy volunteers. The AUC (comparison of those deemed to have glaucoma and borderline disease vs. normal) was 0.84 (95% confidence interval [CI], 0.74 to 0.92) for the cup-to-disc ratio and 0.95 (95% CI, 0.80 to 0.98) for the Disc Damage Likelihood Score, suggesting that the Disc Damage Likelihood Score is a more effective means of discriminating people with and without disease. The diagnostic accuracy of cup-to-disc ratio measurement from the Francis et al. (2011) study is described in the section on FDT C-20 perimetry.				
RNFL Photography **Evidence From Burr et al., 2007.** (4 studies) common cutoff point was diffuse and/or localized defect observed on RNFL photographs. Pooled diagnostic OR 23.1 (95% CI, 4.41 to 123.50) Pooled sensitivity: 75 percent Pooled specificity: 88 percent **Evidence From Primary Studies.** Two studies examined the accuracy of RNFL photography. Hong et al. (2007) analyzed RNFL photographs of 72 glaucoma and 48 healthy	No new studies were identified.	None found	5 reviewers stated the conclusion is still supported by the evidence. 2 reviewers said they did not know.	Original conclusion is still valid and this portion of the original report does not need updating.

Conclusions From CER Executive Summary	RAND Literature Search	FDA/Health Canada/MHRA (UK)	Expert Opinion EPC Investigator Other Experts	Conclusion from SCEPC
participants. Results showed the RNFL defect score II, with an AUC of 0.75 ($p < 0.001$), was the best parameter for discriminating early glaucoma from healthy eyes (sensitivity, 58.3 percent; specificity, 95.8 percent). Medeiros et al. (2004) compared RNFL photography with the GDx with VCC in 42 participants with OAG, 32 persons suspected of having OAG, and 40 healthy volunteers. The sensitivities of the global RNFL score were 36 and 81 percent, respectively, for fixed specificities of 95 and 80 percent. At a fixed specificity of 95 percent, the sensitivity of the Nerve Fiber Indicator was 71 percent versus the 36 percent reported above for red-free photos. Overall, the global RNFL score determined from red-free photos did not perform as well as scanning laser polarimetry. The AUC was 0.91 for the GDx with VCC Nerve Fiber Indicator versus 0.84 for the global RNFL score.				
Scanning Laser Polarimetry (GDx) **Evidence From Primary Studies.** Twenty-seven studies included an investigation of the GDx with VCC. The aim 12 of eight studies was to discriminate early glaucoma from no disease. In the studies that focused on early OAG, the range of sensitivity across all comparisons and cutoffs for the most frequently reported parameter—Temporal, Superior,	Two studies were identified that compared laser polarimetry to other methods for screening: see references 12 and 16 below for new studies pertaining to conclusions for **Direct Comparisons of Candidate Tests**	None found	5 reviewers stated the conclusion is still supported by the evidence. 2 reviewers said they did not know.	Original conclusion is still valid and this portion of the original report does not need updating.

Conclusions From CER Executive Summary	RAND Literature Search	FDA/Health Canada/MHRA (UK)	Expert Opinion EPC Investigator Other Experts	Conclusion from SCEPC
Nasal, Inferior, Temporal average—was 29.8 to 81.63 percent. Specificity was fixed at 80, 90, or 95 percent in three studies, and the lowest reported specificity was 66.36 percent. The range in sensitivity for the nerve fiber indicator parameter across all comparisons and cutoffs was from 28.3 to 93.3 percent. The lowest specificity reported was 52.9 percent or was fixed at 80, 90, or 95 percent. Three studies examined the GDx with enhanced corneal compensation (ECC). The sample sizes of the included studies ranged from 63 to 92 glaucoma participants and 41 to 95 healthy volunteers. Medeiros et al. (2007) compared the AUCs for GDx with VCC and GDx with ECC, and reported that GDx with ECC performed significantly better than GDx with VCC for the parameters Temporal, Superior, Nasal, Inferior, Temporal average, Superior average, and Inferior average (p = <0.01).Sehi et al. (2007) and Mai et al. (2007) concurred with Medeiros et al. (2007) that imaging with ECC appears to improve the ability to diagnose OAG.				
<u>Tests of Optic Nerve Function</u> ***FDT (C-20-1) Perimetry*** **Evidence From Burr et al., 2007.** The pooled sensitivity and specificity results for the three studies that included FDT (C-20-1) perimetry and the common diagnostic criterion of	No new studies were identified.	None found	5 reviewers stated the conclusion is still supported by the evidence. 1 reviewer cited potentially relevant studies. 1 reviewer said s/he did	Original conclusion is still valid and this portion of the original report does not need updating.

Conclusions From CER Executive Summary	RAND Literature Search	FDA/Health Canada/MHRA (UK)	Expert Opinion EPC Investigator Other Experts	Conclusion from SCEPC
one abnormal test point were high (92 and 94 percent, respectively). **Evidence From Primary Studies.** Four studies discussed the accuracy of FDT C-20 perimetry. Pueyo et al. (2009) enrolled 130 participants with ocular hypertension and 48 healthy volunteers. Using a cutoff of a cluster of at least four points with a sensitivity outside 95 percent normal limits, or three points outside 98 percent normal limits, or at least one point outside 99 percent normal limits, investigators determined the sensitivity of FDT to be 31.25 percent and its specificity 72.9 percent among the subset of 32 participants with glaucomatous optic neuropathy (of the 130 with ocular hypertension). The investigators concluded that FDT might not be an ideal test for participants with early defects. Salim et al. (2009) enrolled 35 participants with known OAG and 35 age- and sex-matched controls with no evidence of glaucoma. Investigators used FDT, noncontact tonometry, and a questionnaire individually and in all possible combinations to determine the accuracy of single and combination tests. FDT's sensitivity was 58.1 percent and its specificity was 98.6 percent. Overall, FDT was determined to be the best among the candidate single and combination tests in the study, despite fair sensitivity for			not know.	

Conclusions From CER Executive Summary	RAND Literature Search	FDA/Health Canada/MHRA (UK)	Expert Opinion EPC Investigator Other Experts	Conclusion from SCEPC
detecting OAG. Pierre-Filho et al. (2006) enrolled glaucoma patients who had never experienced perimetry prior to the study. The investigators reported that 21 (32.8 percent) of the 64 participants with glaucoma were identified as having early disease, but data were not provided for this subgroup. Sensitivity and specificity were 85.9 and 73.6 percent, respectively, for the presence of at least one abnormal location and 82.8 and 83 percent, respectively, for two or more abnormal locations, regardless of severity. Francis et al. (2011) conducted population-based screening of 6,082 Latinos age 40 years and older as part of the Los Angeles Latino Eye Study (LALES) to determine the diagnostic accuracy of candidate screening tests performed alone or in combination. Participants completed Humphrey Visual Field testing in addition to FDT C-20-1, GAT, and central corneal thickness and cup-to-disc ratio measurements. Diagnostic test accuracy outcomes were assessed for the general population as well as high-risk subgroups, defined as persons who were 65 years and older, those with a family history of glaucoma, and persons with diabetes. Of the 6,082 participants screened, 4.7 percent (286) were diagnosed as having OAG. Based on three glaucoma diagnosis				

Conclusions From CER Executive Summary	RAND Literature Search	FDA/Health Canada/MHRA (UK)	Expert Opinion EPC Investigator Other Experts	Conclusion from SCEPC
definitions (glaucomatous optic nerve appearance, glaucomatous visual field loss, glaucomatous optic nerve and visual field loss), the test parameters vertical cup-to-disc ratio ≥ 0.8 and Humphrey Visual Field (HVF) false negatives ≥ 33 percent had the highest specificity, regardless of the definition of glaucoma (98 percent). HVF mean deviation < 5 percent had the highest sensitivity (78 percent) using the definition of optic nerve defects only, while the HVF glaucoma hemifield test had the highest sensitivity under the other two definitions (90 percent for glaucomatous visual field loss and 90 percent for both field loss and optic nerve damage). Specific results for the FDT C-20-1 were as follows (sensitivity/specificity, definition of glaucoma): 59 percent/ 79 percent, glaucomatous optic nerve appearance only; 68 percent/80 percent, glaucomatous visual field loss only; 67 percent/79 percent, both glaucomatous optic nerve appearance and visual field loss. The investigators reported similar results when high-risk subgroups were analyzed and concluded that "these results suggest that screening of high-risk groups based on these criteria may not improve over screening of the general population over age 40."				
<u>FDT (C-20-5) Perimetry</u> **Evidence From Burr et al., 2007.** Five studies of FDT (C-20-5) with significant heterogeneity using the	No new studies were identified.	None found	5 reviewers stated the conclusion is still supported by the evidence.	

Conclusions From CER Executive Summary	RAND Literature Search	FDA/Health Canada/MHRA (UK)	Expert Opinion EPC Investigator Other Experts	Conclusion from SCEPC
common cutoff point of one abnormal test point were included. The range of sensitivity was 7 to 100 percent; the specificity range was 55 to 89 percent. ***FDT 24-2 Perimetry*** **Evidence From Primary Studies.** Five studies examined the diagnostic accuracy of FDT 24-2 threshold tests using the Humphrey Matrix Perimeter. All studies included participants with known glaucoma and healthy volunteers, and we judged these studies to be at high risk of spectrum bias. The range of sample size was 25 to 174 glaucomatous eyes and 15 to 164 healthy eyes. Sensitivities and specificities were reported for the parameters mean deviation, pattern standard deviation, and glaucoma hemifield test outside of normal limits. There was appreciable heterogeneity in the estimates of sensitivity at 80 percent, 90 percent, and 95 percent specificity that may be attributed to a number of factors, including different patient populations and variations in cutoff points. The sensitivity was 55 percent for the mean deviation and 94 percent at 80 percent fixed specificity. Tafreshi et al. (2009) and Leeprechanon et al. (2007) reported 39 and 87 percent at 90 percent fixed specificity, and 32 and 82 percent at fixed 95 percent specificity, respectively. Sensitivity and specificity for pattern standard of			2 reviewers said they did not know.	

Conclusions From CER Executive Summary	RAND Literature Search	FDA/Health Canada/MHRA (UK)	Expert Opinion EPC Investigator Other Experts	Conclusion from SCEPC
deviation (PSD) and glaucoma hemifield test are reported with their cutoff points in the evidence tables in Appendix C of the full report. Bagga et al. (2006) and Burgansky-Eliash et al. (2007) reported the AUC for the mean deviation parameter (0.69 for both studies with $p < 0.04$ and 95% CI, 0.564 to 0.815, respectively). The AUCs for PSD were 0.66 ($p = 0.09$) and 0.733 (95% CI, 0.618 to 0.848).				
FDT 30-2 Perimetry **Evidence From Primary Studies.** Two studies discussed the detection of early glaucoma using the FDT 30-2 threshold test with the Humphrey Matrix Perimeter. Both Hong, Chung, Hong, et al. (2007) and Hong, Ahn, Ha, et al. (2007) enrolled OAG participants with early visual field loss and healthy controls. The mean deviation and PSD were judged to be good parameters for distinguishing between eyes with early disease and eyes with no known defects. The mean deviations were 0.795 and 0.750 and the PSDs were 0.808 and 0.934 for Hong, Chung, Hong, et al. and Hong, Ahn, Ha, et al., respectively. Both study groups, however, determined that the best parameter for distinguishing eyes with early glaucoma from healthy eyes was the number of points that have p less than 5 percent in the pattern deviation plot, with an AUC of 0.985 (95% CI, 0.943	No new studies were identified.	None found	4 reviewers stated the conclusion is still supported by the evidence. 3 reviewers said they did not know.	Original conclusion is still valid and this portion of the original report does not need updating.

Conclusions From CER Executive Summary	RAND Literature Search	FDA/Health Canada/MHRA (UK)	Expert Opinion EPC Investigator Other Experts	Conclusion from SCEPC
to 0.998) in Hong, Chung, Hong, et al. and 0.990 ($p < 0.001$) in Hong, Ahn, Ha, et al.				
FDT N-30 Perimetry **Evidence From Primary Studies.** Four studies examined the accuracy of the FDT N-30 threshold test. Zeppieri et al. (2010) focused on the detection of early glaucoma among a sample of 75 participants with OAG, 87 with ocular hypertension, 67 with glaucomatous optic neuropathy, and 90 healthy volunteers. At the best cutoff of less than -0.78, the sensitivity of the mean deviation parameter was 61.3 percent and the specificity was 73.7 percent for distinguishing early OAG from healthy eyes. At the best cutoff of greater than 3.89, the sensitivity of the PSD was 76.0 percent and the specificity was 87.8 percent. Salvetat et al. (2010) focused on the detection of early disease among a sample of 52 participants with early OAG and 53 healthy volunteers. The sensitivity of mean deviation for distinguishing early OAG from healthy eyes at the best cutoff (less than -1.12) was 67 percent and the specificity was 74 percent. At the best cutoff of greater than 3.97, the sensitivity of the parameter PSD was 96 percent and the specificity was 85 percent.	No new studies were identified.	None found	5 reviewers stated the conclusion is still supported by the evidence. 2 reviewers said they did not know.	Original conclusion is still valid and this portion of the original report does not need updating.
Goldmann Applanation Tonometry **Evidence From Burr et al., 2007.** At	One new study that assessed the use of iCare rebound tonometry compared with	None found	5 reviewers stated the conclusion is still	Original conclusion is still valid and this

Conclusions From CER Executive Summary	RAND Literature Search	FDA/Health Canada/MHRA (UK)	Expert Opinion EPC Investigator Other Experts	Conclusion from SCEPC
the common cutoff point of IOP greater than 20.5-22 mm Hg, nine studies with significant heterogeneity reported sensitivity in the range of 10 to 90 percent and specificity in the range of 81 to 99 percent. **Evidence From Primary Studies.** Two studies included examination of GAT. Bagga et al. (2006) compared the ability of various tests of structure and function to discriminate healthy eyes (n = 22) from eyes with known glaucomatous optic neuropathy (n = 25). The AUC for IOP, as measured by GAT, was 0.66 (p = 0.05). The methods of the Francis et al. (2011) study (LALES) are discussed in the FDT C-20 section of the full report. The specific sensitivity and specificity values for GAT using a cutoff of ≥ 21 mm Hg for the three definitions of glaucoma were as follows (sensitivity/specificity, definition of glaucoma): 21 percent/97 percent, glaucomatous optic nerve appearance only; 23 percent/97 percent, glaucomatous visual field loss only; 24 percent/97 percent, both glaucomatous optic nerve appearance and visual field loss.	GAT[14] found that iCare recorded slightly higher values than GAT but had high sensitivity and specificity and can be used by paramedical personnel for screening. Another study found that using a central corneal thickness(CCT)-based formula to correct GAT measures resulted in comparatively poorer agreement with Pascal dynamic contour tonometry than unadjusted GAT,[6] especially with thicker corneas; however the authors concluded that this correction might be useful in population screening.		supported by the evidence. 1 reviewer cited potentially relevant studies on diurnal variation in IOP but didn't think they are relevant to glaucoma. 1 reviewer said s/he did not know.	portion of the original report does not need updating.
<u>Humphrey Visual Field Analyzer</u> **Evidence From Primary Studies.** Ten studies examined the diagnostic accuracy of the HFA. Of these, six examined HFA Short Wavelength Automated Perimetry; two tested	No new studies were identified.	None found	4 reviewers stated the conclusion is still supported by the evidence. 3 reviewers said they did not know.	Original conclusion is still valid and this portion of the original report does not need updating.

Conclusions From CER Executive Summary	RAND Literature Search	FDA/Health Canada/MHRA (UK)	Expert Opinion EPC Investigator Other Experts	Conclusion from SCEPC
HFA-SAP, SAP-SITA, and HFA SAP-Full Threshold (FT); four examined HFA-SITA-Standard; and one tested the HFA SITA-Fast protocol. The HFA Short Wavelength Automated Perimetry testing protocol (the most frequently reported) included 25 to 286 participants with glaucoma and 22 to 289 healthy volunteers across the six included studies. Sensitivity across all comparisons and cutoffs for the mean deviation ranged from 25.9 to 83 percent. Specificity ranged from 80 to 95.2 percent. Cutoff points ranged from -5.42 to -11.06 dB.				
Noncontact Tonometry **Evidence From Burr et al., 2007.** (1 study) Sensitivity: 92 percent Specificity 92 percent (using the criterion of IOP greater than 21 mm Hg). **Evidence From Primary Studies.** Salim et al. (2009) included noncontact tonometry, individually and in all possible combinations, with other measures of structure and function to determine the accuracy of single and combination tests. IOP, as measured by noncontact tonometry, was found not to be a very sensitive test for detecting glaucoma (sensitivity 22.1 percent). The investigators acknowledge that use of topical medications by the glaucoma participants could limit the ability to	No new studies were identified.	None found	4 reviewers stated the conclusion is still supported by the evidence. 3 reviewers said they did not know, although 1 of the 3 mentioned new literature on corneal hysteresis that might be relevant.	Original conclusion is still valid and this portion of the original report does not need updating.

Conclusions From CER Executive Summary	RAND Literature Search	FDA/Health Canada/MHRA (UK)	Expert Opinion EPC Investigator Other Experts	Conclusion from SCEPC
identify those with disease.				
Oculokinetic Perimetry **Evidence From Burr et al., 2007.** Four studies were included that examined the diagnostic accuracy of oculokinetic perimetry. The common criterion varied in description, but is best described as one or more points missing. The odds of a positive test were 57 times as high (DOR, 57.54) for those with glaucoma as for those without glaucoma (95% CrI, 4.42 to 1585.00). The pooled sensitivity and specificity were 86 and 90 percent, respectively.	No new studies were identified.	None found	4 reviewers stated the conclusion is still supported by the evidence. 3 reviewers said they did not know.	Original conclusion is still valid and this portion of the original report does not need updating.
SAP Suprathreshold Test **Evidence From Burr et al., 2007.** Nine studies, including the Baltimore Eye Survey and the Blue Mountains Eye Study, examined the SAP suprathreshold test. Although the sensitivity and specificity were similar for the Baltimore and Blue Mountains studies, there was significant heterogeneity among the included studies. The range in sensitivity was 25 to 90 percent; the range in specificity was 67 to 96 percent.	No new studies were identified.	None found	5 reviewers stated the conclusion is still supported by the evidence. 2 reviewers said they did not know.	Original conclusion is still valid and this portion of the original report does not need updating.
SAP Threshold Test **Evidence From Burr et al., 2007.** Among the five studies analyzed for SAP threshold, both Humphrey 30-2 and 24-2 threshold and Octopus 500 were evaluated. The pooled sensitivity was 88 percent, and specificity was 80	No new studies were identified.	None found	6 reviewers stated the conclusion is still supported by the evidence, including 1 who said that SAP and SITA are falling out of favor.	Original conclusion is still valid and this portion of the original report does not need updating.

Conclusions From CER Executive Summary	RAND Literature Search	FDA/Health Canada/MHRA (UK)	Expert Opinion EPC Investigator Other Experts	Conclusion from SCEPC
percent for the common cutoff point. (The definition of the common cutoff point differed by included study, but is defined in Burr et al.)			1 reviewer said s/he did not know.	
Tendency-Oriented Perimetry **Evidence From Primary Studies.** Pierre-Filho et al. (2006) compared frequency doubling technology), tendency-oriented perimetry using the Octopus 301 G1-TOP program, SITA Standard, and SITA Fast in 117 eyes (64 with glaucoma and 53 healthy eyes). The Octopus 301 perimeter test was considered abnormal under two conditions: when the mean defect was "> 2dB and/or the loss variance > 6 dB (TOP 1), and…there were at least seven points (three of them contiguous) with a reduction in sensitivity ≥ 5 dB in the corrected comparisons graphic (TOP 2)."92 The sensitivity using definition TOP 1 was 87.5 percent (95% CI, 76.3 to 94.1) and the specificity was 56.6 percent (95% CI, 42.4 to 69.9). With definition TOP 2, the sensitivity was 89.1 percent (95% CI, 78.2 to 95.1) and the specificity was 62.3 percent (95% CI, 47.9 to 74.9).	No new studies were identified.	None found	5 reviewers stated the conclusion is still supported by the evidence. 2 reviewers said they did not know.	Original conclusion is still valid and this portion of the original report does not need updating.
Direct Comparisons of Candidate Tests **Evidence From Burr et al., 2007** Six studies included comparisons of SAP with optic disc photography, HRT II, FDT, and/or GAT. Burr et al. concluded that sensitivity results at the common cutoff point for each test	One study compared the use of TD-OCT to measure renal nerve fiber layer (RNFL) thickness with GDx-VCC scanning laser polarimetry.[12] At average RNFL thickness, OCT had the higher diagnostic accuracy for detecting early defects (AUC 0.785). GDx-VCC had an AUC of 0.758 at temporal-superior-nasal-	None found	3 reviewers stated the conclusion is still supported by the evidence. 2 reviewers said they did not know 2 reviewers cited potentially relevant	Original conclusion is still valid and this portion of the original report does not need updating.

Conclusions From CER Executive Summary	RAND Literature Search	FDA/Health Canada/MHRA (UK)	Expert Opinion EPC Investigator Other Experts	Conclusion from SCEPC
revealed that SAP performed better than GAT. One of the two studies that addressed the comparison of SAP to GAT reported estimates of sensitivity of 89 percent and 3 to 14 percent, respectively. Specificity values were 73 percent for SAP and 98 to 99 percent for GAT. Burr et al. also concluded that SAP was similar to HRT II. The sensitivity of SAP was 72 percent and the sensitivity of HRT II was 69 percent in one of the two included studies; the specificity for both tests was 95 percent. There was one included study in which the investigators compared SAP with optic disc photography. Optic disc photographs had a similar sensitivity (73 to 77 percent) and specificity (59 to 62 percent) to SAP (sensitivity, 50 to 71 percent; specificity, 58 to 83 percent). In the two studies that included comparisons of SAP with FDT, one study reported similar sensitivity estimates (SAP, 63 to 90 percent; FDT C-20-5, 68 to 84 percent) and similar specificity values (SAP, 58 to 74 percent; FDT C-20-5, 55 to 76 percent). Based on analyses of the common criterion for each test, test accuracy, combination tests, tests for glaucoma at specific stages, and direct and indirect comparisons of tests, Burr et al. (2007) concluded that optic disc photography, HRT II, FDT, SAP, and GAT were candidates for use in a	inferior-temporal; thus both procedures had similar diagnostic accuracy for identifying early defects in OHT patients. Another study compared FD-OCT of RNFL thickness with that of both GDx-VCC and ECC scanning laser polarimetry.[16] All 3 methods had high specificity but RNFLT damage, OCT was significantly more sensitive than scanning polarimetry. 87% of cases were diagnosed by all 3 methods.		studies either comparing tests or assessing combinations of tests.	

Conclusions From CER Executive Summary	RAND Literature Search	FDA/Health Canada/MHRA (UK)	Expert Opinion EPC Investigator Other Experts	Conclusion from SCEPC
screening-based program.				
Conclusion Based on the Burr et al. (2007) findings, standard automated perimetry was compared with other tests available at that time. SAP had higher sensitivity than Goldmann tonometry, similar sensitivity to HRT, and lower sensitivity than disc photos or FDT. In terms of specificity, SAP performed better than disc photos and FDT, similar to HRT, and worse than Goldmann tonometry. **Evidence from Primary Studies:** Several additional studies assessed the performance of glaucoma screening tests not included in the Burr et al. review. The studies included newer imaging (GDx, HRT III, OCT) and functional (Short Wavelength Automated Perimetry, new FDT patterns) technologies. However, despite improvements in the technology, it is still not clear that there is any one test or combination of tests suitable for use in glaucoma screening in the general population. Significant barriers to identifying and characterizing potential glaucoma screening tests remain. These barriers include the lack of a definitive diagnostic reference standard for glaucoma and heterogeneity in the design and conduct of the studies. Because of these barriers, the ranges of sensitivities, specificities, and	No new studies were identified.	None found	3 reviewers stated the conclusion is still supported by the evidence. 3 reviewers said they did not know. 1 reviewer did not respond.	Original conclusion is still valid and this portion of the original report does not need updating.

Conclusions From CER Executive Summary	RAND Literature Search	FDA/Health Canada/MHRA (UK)	Expert Opinion EPC Investigator Other Experts	Conclusion from SCEPC
AUCs are large and prevent a coherent synthesis.				
KQ4a: Does a screening-based program for open-angle glaucoma lead to reductions in intraocular pressure when compared with no screening program? KQ4b: How does intraocular pressure vary when comparing different screening-based programs for open-angle glaucoma?				
No studies were identified that addressed whether participation in an OAG screening-based program leads to reductions in IOP when compared with no screening or another screening-based program.	No new studies were identified.	None found	5 reviewers stated the conclusion is still supported by the evidence. 2 reviewers said they did not know.	Original conclusion is still valid and this portion of the original report does not need updating.
KQ5a: Does a screening-based program lead to a slowing of the progression of optic nerve damage and visual field loss when compared with no screening program? KQ5b: How do optic nerve damage and visual field loss vary when comparing different screening-based programs for open-angle glaucoma?				
Evidence From Systematic Reviews A systematic review of randomized trials of screening modalities for OAG compared with no screening (including opportunistic case finding and referral) (Hatt et al., 2006). There were no restrictions on included populations. The primary outcome of interest was the prevalence of visual field loss, defined as the proportion of participants with a prespecified severity of visual field loss diagnosed by either manual or automated field assessment. Other primary outcomes included the prevalence of optic nerve damage and visual impairment. Electronic searches of five databases, including MEDLINE and CENTRAL, were conducted in 2006 and again in January 2009, but none of the studies that were identified were eligible for	No new studies were identified.	None found	4 reviewers stated the conclusion is still supported by the evidence. 3 reviewers said they did not know.	Original conclusion is still valid and this portion of the original report does not need updating.

Conclusions From CER Executive Summary	RAND Literature Search	FDA/Health Canada/MHRA (UK)	Expert Opinion EPC Investigator Other Experts	Conclusion from SCEPC
inclusion. The review authors acknowledged that RCTs require lengthy followup and are predicated on identifying appropriate candidate tests that may be incorporated into a screening-based program. **Detailed Analysis of Primary Studies** We did not identify any study that addressed whether participation in an OAG screening-based program leads to reductions in visual field loss or optic nerve damage when compared with no screening or another screening-based program.				
KQ 6: What are the harms associated with screening for open angle glaucoma?				
We did not identify any study addressing the harms associated with screening for OAG.	No new studies were identified.	None found	2 reviewers stated the conclusion is still supported by the evidence. 5 reviewers did not respond.	Original conclusion is still valid and this portion of the original report does not need updating.
Are there new data that could inform the key questions that might not be addressed in the conclusions? One reviewer stated that the concept of combining structure and functional tests in the evaluation may provide information. There may be new studies looking at this particular area. One review mentioned a new detection method: the use of OCT to measure the ganglion cell complex. One reviewer suggested 2 new articles on screening, one a non-systematic review and one on using a Delphi process to design a screening program. The article on using the Delphi method[17] reported the results of a 38-member panel that agreed with a scenario that beginning at age 50, screening should occur using two or three tests in varying combinations (including tonometry (IOP), measures of visual function (e.g., FDT), measures of optic nerve damage (e.g., HRT, OCT, and cup/disc ratio), organized in a community setting with testing being conducted by trained paraprofessionals. An alternative was a glaucoma risk score based on a questionnaire.				

Legend: AUC Area under the receiver operating characteristic curve; DDLS Disk Damage Likelihood Scale; FD Fourier Domain; FDT frequency doubling technology; GPS glaucoma probability score; HRT Heidelberg Retina Tomograph; MGSCP Mobile Glaucoma Screening Clinic Project; LCDR Linear Cup to Disc Ratio; MRA Moorfields regression analysis; NFL nerve fiber layer; OAG open-angle glaucoma; OCT optical coherence tomography; OHT ocular hypertension; ONH Optic Nerve Head; RAPD Relative Afferent Pupillary Defect; RNFL Renal Nerve Fiber Layer; SCEPC Southern California Evidence-based Practice Center; SD spectral domain; TD time domain

References

1. Ervin AM, Boland MV, Myrowitz EH, et al. Screening for Glaucoma: Comparative Effectiveness. Comparative Effectiveness Review No. 59. (Prepared by the Johns Hopkins University Evidence-based Practice Center under Contract No. 290-2007-10061.) AHRQ Publication No. 12-EHC037-EF. Rockville, MD: Agency for Healthcare Research and Quality. www.effectivehealthcare.ahrq.gov/reports/final.cfm. April 2012 AprPMID 22649799.

2. Shekelle PG, Newberry SJ, Maglione M, et al. Assessment of the Need to Update Comparative Effectiveness Reviews: Report of an Initial Rapid Program Assessment (2005-2009) (Prepared by the Southern California Evidence-based Practice Center). Rockville, MD: Agency for Healthcare Research and Quality; October 2009.

3. Shojania KG, Sampson M, Ansari MT, et al. How quickly do systematic reviews go out of date? A survival analysis. Ann Intern Med. 2007 Aug 21;147(4):224-33. PMID 17638714.

4. Chen HY, Chang YC, Wang IJ, et al. Comparison of Glaucoma Diagnoses Using Stratus and Cirrus Optical Coherence Tomography in Different Glaucoma Types in a Chinese Population. J Glaucoma. 2012 May 16PMID 22595933.

5. Hennessy AL, Katz J, Ramakrishnan R, et al. The utility of relative afferent pupillary defect as a screening tool for glaucoma: prospective examination of a large population-based study in a south Indian population. Br J Ophthalmol. 2011 Sep;95(9):1203-6. PMID 21349935.

6. Park SJ, Ang GS, Nicholas S, et al. The effect of thin, thick, and normal corneas on Goldmann intraocular pressure measurements and correction formulae in individual eyes. Ophthalmology. 2012 Mar;119(3):443-9. PMID 22035576.

7. Rao HL, Kumbar T, Addepalli UK, et al. Effect of spectrum bias on the diagnostic accuracy of spectral-domain optical coherence tomography in glaucoma. Invest Ophthalmol Vis Sci. 2012 Feb;53(2):1058-65. PMID 22266520.

8. Um TW, Sung KR, Wollstein G, et al. Asymmetry in hemifield macular thickness as an early indicator of glaucomatous change. Invest Ophthalmol Vis Sci. 2012 Mar;53(3):1139-44. PMID 22247461.

9. Kamdeu Fansi AA, Agoumi Y, Harasymowycz PJ. Screening for glaucoma with Moorfields regression analysis and glaucoma probability score in confocal scanning laser ophthalmoscopy. Can J Ophthalmol. 2011 Jun;46(3):254-60. PMID 21784211.

10. Kamdeu Fansi AA, Boisjoly H, Chagnon M, et al. Combining rim area to disc area asymmetry ratio and Moorfields regression analysis of confocal scanning laser ophthalmoscopy for glaucoma screening. Can J Ophthalmol. 2011 Jun;46(3):261-6. PMID 21784212.

11. Kaushik S, Singh Pandav S, Ichhpujani P, et al. Retinal nerve fiber layer measurement and diagnostic capability of spectral-domain versus time-domain optical coherence tomography. Eur J Ophthalmol. 2011 Sep-Oct;21(5):566-72. PMID 21279977.

12. Pablo LE, Ferreras A, Schlottmann PG. Retinal nerve fibre layer evaluation in ocular hypertensive eyes using optical coherence tomography and scanning laser polarimetry in the diagnosis of early glaucomatous defects. Br J Ophthalmol. 2011 Jan;95(1):51-5. PMID 20576777.

13. Ramdas WD, Wolfs RC, Hofman A, et al. Heidelberg Retina Tomograph (HRT3) in population-based epidemiology: normative values and criteria for glaucomatous optic neuropathy. Ophthalmic Epidemiol. 2011 Oct;18(5):198-210. PMID 21961509.

14. Scuderi GL, Cascone NC, Regine F, et al. Validity and limits of the rebound tonometer (ICare(R)): clinical study. Eur J Ophthalmol. 2011 May-Jun;21(3):251-7. PMID 20853257.

15. Pomorska M, Krzyzanowska-Berkowska P, Misiuk-Hojlo M, et al. Application of optical coherence tomography in glaucoma suspect eyes. Clin Exp Optom. 2012 Jan;95(1):78-88. PMID 21981362.

16. Garas A, Vargha P, Hollo G. Comparison of diagnostic accuracy of the RTVue Fourier-domain OCT and the GDx-VCC/ECC polarimeter to detect glaucoma. Eur J Ophthalmol. 2012 Jan-Feb;22(1):45-54. PMID 22167542.

17. Campbell SE, Azuara-Blanco A, Campbell MK, et al. Developing the specifications of an Open Angle Glaucoma screening intervention in the United Kingdom: a Delphi approach. BMC Health Serv Res. 2012;12:447. PMID 23216983.

18. Rao HL, Leite MT, Weinreb RN, et al. Effect of disease severity and optic disc size on diagnostic accuracy of RTVue spectral domain optical coherence tomograph in glaucoma. Invest Ophthalmol Vis Sci. 2011 Mar;52(3):1290-6. PMID 20811060.

19. Charalel RA, Lin HS, Singh K. Glaucoma Screening Using Relative Afferent Pupillary Defect. J Glaucoma. 2013 Jan 3PMID 23296370.

20. Bengtsson B, Andersson S, Heijl A. Performance of time-domain and spectral-domain Optical Coherence Tomography for glaucoma screening. Acta Ophthalmol. 2012 Jun;90(4):310-5. PMID 20946342.

21. Healey PR, Lee AJ, Aung T, et al. Diagnostic accuracy of the Heidelberg Retina Tomograph for glaucoma a population-based assessment. Ophthalmology. 2010 Sep;117(9):1667-73. PMID 20816247.

22. Lisboa R, Leite MT, Zangwill LM, et al. Diagnosing preperimetric glaucoma with spectral domain optical coherence tomography. Ophthalmology. 2012 Nov;119(11):2261-9. PMID 22883689.

Appendices

Appendix A: Search Methodology

Appendix B: Evidence Tables

Appendix C: Questionnaire Matrix

Appendix A. Search Methodology

Searched 12/26/2012

Update search covering 8/01/2011 – present
("Ocular Hypertension"[mh] OR "ocular hypertension"[tiab] OR "Intraocular Pressure"[mh] OR "intraocular pressure"[tiab] OR “glaucoma, open-angle” [mh] OR “Open angle glaucoma” [tiab] OR “low tension glaucoma” [tiab] OR “normal tension glaucoma” [tiab] OR “pseudoexfoliative glaucoma” [tiab] OR “pseudoexfoliative syndrome” [tiab])**AND** (screening[tiab] OR "early diagnosis"[mh] OR "tomography, optical coherence"[mh] OR tomography[tiab] OR OCT OR "tonometry, ocular" [mh] OR perimetry[tiab] OR HRT[tiab] OR "Heidelberg retina tomograph" [tiab] OR "scanning laser polarimetry"[tiab]) **AND** ("Cohort Studies"[Mesh] OR "Case-Control Studies"[Mesh] OR randomized controlled trial [pt] OR controlled clinical trial [pt] OR randomized [tiab] OR placebo [tiab] OR drug therapy [sh] OR randomly [tiab] OR trial [tiab] OR groups [tiab]) NOT (animals [mh] NOT humans [mh])**AND** Journal of Glaucoma[journal] OR American Journal of Ophthalmology[journal] OR ophthalmology[journal] OR investigative ophthalmology & visual science[journal] OR archives of ophthalmology[journal] OR “the British Journal of Ophtalmology”[journal]
AND
("Annals of internal medicine"[Journal] OR "bmj"[Journal] OR "jama"[Journal] OR "lancet"[Journal] OR "The New England journal of medicine"[Journal])

Results: 248

Appendix B. Evidence Table

Study **Region** **Study Design**	**Populations (n, mean age, gender number, %)**	**Inclusion criteria** **Exclusion criteria**	**Reference standard(s)** **Tests carried out and interpreted by**	**Device**	**Findings**
Screening Predictive Value					
Swinging Flashlight RAPD					
Charalel 2013[19] US Population-based cross-sectional study	107 subjects, including those with and without glaucoma and glaucoma suspects	Inclusion age≥18 ability to read consent forms	History and ophthalmic examination	Swinging flashlight to assess relative afferent pupillary defect (RAPD)	RAPD screening with swinging flashlight and neutral density filters was moderately sensitive and strongly specific (OR 9.71, sensitivity 66.7%, specificity 82.9%). Exclusion of prior cataract surgery patients improved sensitivity, specificity , and PPV
Hennessy 2011[5] India Prospective population-based study	18,386 participants in the Aravind Comprehensive Eye Survey (ACES)	Inclusion Residence in catchment areas	Trained ophthalmic technicians	Swinging flashlight to assess RAPD	In the rural village clinic setting, RAPD was a poor screening tool for glaucoma: Sensitivity for glaucoma: 0% Specificity 99.8% Sensitivity of household screening for APD was 1.7%; sensitivity of

Study **Region** **Study Design**	**Populations (n, mean age, gender number, %)**	**Inclusion criteria** **Exclusion criteria**	**Reference standard(s)** **Tests carried out and interpreted by**	**Device**	**Findings**
					hospital comprehensive exam for APD was 9.1%, specificity 99.3%
IOP Measurement					
Scuderi 2010[14] Italy Non-randomized trial	98 patients: 59 women (mean age 53.5±19.4) 38 men (mean age 53.1±17.8)	age≥30 Suspected or dx glaucoma Excluded: infectious or inflammatory eye pathology	GAT (>21 mm Hg) 2 trained examiners masked to IOP data	iCare rebound tonometer	iCare tonometry recorded slightly higher values than GAT. Sensitivity 0.90 and specificity 0.95 iCare is comparable with GAT and can be used by paramedical personnel during population screening
Park 2012[6] New Zealand Retrospective cross-sectional case series	289 patients attending a specialist glaucoma practice, normal subjects, and patients with confirmed glaucomatous optic neuropathy	Exclusion criteria: corneal diseases, connective tissue disorders, secondary glaucoma, previous ocular trauma	Pascal dynamic contour tonometer (PDCT) to measure IOP DX based on glaucomatous appearance of optic disc, visual field defects and verification by glaucoma specialist	Central cornea thickness (CCT)-based correction formulae for GAT-measured IOP for stratified CCT groups	Adjusting IOP with CT-based formulae resulted in poorer agreement with PDCT IOP than unadjusted GAT IOP. Risk of significant error, especially with thicker corneas. Suggests that CCT may be useful in population analysis but not when applied to

Study **Region** **Study Design**	**Populations (n, mean age, gender number, %)**	**Inclusion criteria** **Exclusion criteria**	**Reference standard(s)** **Tests carried out and interpreted by**	**Device**	**Findings**
					individuals
HRT3					
Healey 2010[21] Blue Mountains Eye Study Population-based cross-sectional study	N=1952, 75.6% of survivors, 10-year follow-up Mean age 73.7 years	Described in an earlier paper Individuals enrolled in longterm screening study	GAT, visual field testing, slit-lamp anterior segment exam	HRT scans using Moorfield's Regression Analysis (MRA)	87.4% of scans had a topography standard deviation ≤40um. Larger TSD associated with older age and OAG. MRA sensitivity was 64.1%, specificity 85.7%, PPV21%, NPV 97.6% for detecting OAG. Including borderline results improved sensitivity to 87.0% but specificity dropped to 70.6%. TSD restriction to <40um improved diagnostic accuracy. Single eye scans overestimated specific and underestimated sensitivity compared with using data from both eyes. The specificity of the MRA was inadequate for use as a screening test but was acceptable in an older

Study Region Study Design	Populations (n, mean age, gender number, %)	Inclusion criteria Exclusion criteria	Reference standard(s) Tests carried out and interpreted by	Device	Findings
					population with acceptable quality scans in most eyes.
Kamdeu Fansi 2011a[9] Canada Population-based, cross-sectional study of proposed screening tools	MGSCP participants N=232 Mean age=61±11 151/81 178 Caucasian/54 African/Caribbean	Inclusion: individuals at high-risk for OAG (Caribbean or African descent, > 50 yoa, + fx hx for OAG	FDT perimetry and ophthalmic exam (gonioscopy, IOP, slit lamp exam, observation of optic disk, NFL, and retina, vertical cup/disk ratio and DDLS, by specialist blinded to HRT results) Ref Std. 1: (normal+possible glaucoma+probable vs. Definitive) Ref Std. 2: (normal+possible vs. probable +definitive)	HRT II/III (scanning laser ophthalmoscopy)	Depending on the reference standard diagnosis and the borderline test positive definition, the sensitivity varied from 36% to 100% and the specificity varied from 60% to 97%; the MRA using HRT3 appears to be more sensitive but less specific than the MRA using HRTII. GPS was less specific but more sensitive, and as a screening method has the advantage of being independent of contour lines.
Kamdeu Fansi 2011b[10] Canada Subjects at high risk for developing POAG Population-based, cross-sectional study of proposed	MGSCP participants N=375 Mean age differed by risk group Approx 1/6 African/Caribbean to Caucasian	Inclusion: individuals at high-risk for OAG (Caribbean or African descent, > 50 yoa, + fx hx for OAG	Reference standards: see Fansi 2011a; Standard opt. exam(see Fansi 2011a), FDT, MRA-HRTIII	HRTIII	In high-risk populations, RADAAR had a higher specificity than MRA in identifying glaucoma. For patients who were borderline on MRA,

Study **Region** **Study Design**	**Populations (n, mean age, gender number, %)**	**Inclusion criteria** **Exclusion criteria**	**Reference standard(s)** **Tests carried out and interpreted by**	**Device**	**Findings**
screening tools					combining MRA with RADAAR increased the specificity
Ramdas 2011[13] Netherlands Rotterdam Study (aim was to establish normative HRT3 values) Prospective cohort study	2516 Consecutive Rotterdam Study Participants Caucasian	age≥55, residence in Rotterdam	Ophthalmic assessment included Hx, Goldmann applanation tonometry, autorefraction, keratometry, ophthalmometry, best-corrected visual acuity and field testing , and ophthalmoscopy and optic nerve head (ONH) fundoscopy with HRT and image net; Visual field screening with Humphrey Field Analyzer HFA II perimetry and HFA24-2 if indicated to measure visual field loss	HRT3 (Linear cup disc ratio[LCDR])	At the required specificity of 97.5%, HRT3 sensitivity adjusted for disc area is low (although greater than VCDR) The HRT3 Glaucoma Probability Score (GPS) and linear discrimination functions had lower sensitivity. Thus adjusted LCDR seems to be most suited to develop criteria for population studies
OCT					
Rao, 2011[18] California Diagnostic Innovations in Glaucoma Study (observational cross-sectional)	62 normal subjects and 136 glaucoma patients of *varying disease severity*	Inclusion: Best corrected visual acuity of 20/40 or better, spherical refraction within ±5.0D, and cylinder correction within ±3.0D and open angles Exclusion: co-existing retinal disease, uveitis, nonglaucomatous	VFI values from standard automated perimetry	RTVue spectral domain(SD) OCT HRT2 used to measure disc size	Diagnostic accuracies of RTVue scanning protocols for glaucoma were significantly influenced by

Study **Region** **Study Design**	**Populations (n, mean age, gender number, %)**	**Inclusion criteria** **Exclusion criteria**	**Reference standard(s)** **Tests carried out and interpreted by**	**Device**	**Findings**
		optic neuropathy			disease severity. Sensitivity of the rim area increased in large optic discs but specificity decreased thus optic disc size did not influence AUC of any of the SDOCT scanning protocols.
Rao 2012[7] India Case control	60 normal subjects (119 eyes), 41 glaucoma suspects (76 eyes), 46 early glaucoma patients (65 eyes)	Inclusion: age≥18 Best corrected visual acuity 20/40 or better, refractive error within ±5.0D and cylinder correction within ±3.0D Exclusion: media opacities, intraocular surgery within prior 6 months, any retinal or neurological diseases other than glaucoma	Not relevant	SD-OCT of optic nerve head, retinal nerve fiber layer, and ganglion cell complex	Effectiveness of most SD-OCT parameters in detecting glaucoma significantly decreased (as measure by AUC) when evaluated against a clinically relevant control group of glaucoma suspects compared with evaluation against group with no glaucoma. Therefore, choice of control group is important when validating dx test.
Um 2012[8] Seoul Korea Prospective cohort study	114 healthy, 103 glaucoma suspects, and 74 glaucomatous eyes	Not reported	Not reported (RFNL)	SD-OCT to measure Hemifield macular Thickness vs. RFNL	Macular hemifield test showed better performance than average cRFNL thickness measurements in

Study **Region** **Study Design**	**Populations (n, mean age, gender number, %)**	**Inclusion criteria** **Exclusion criteria**	**Reference standard(s)** **Tests carried out and interpreted by**	**Device**	**Findings**
					diagnostic sensitivity in patients with *early stage glaucoma*, with similar specificity
Lisboa 2012[22] US (UCSD) Diagnostic Innovations in Glaucoma Study (Cohort study)	134 eyes of 88 glaucoma suspects	Inclusion: Best corrected visual acuity 20/40 or better, refractive error within ±5.0D and cylinder correction within ±3.0D, open angle with gonioscopy Exclusion: co-existing retinal disease, uveitis, nonglaucomatous optic disc neuropathy	Suspicious appearance of optic nerve from cross sectional evaluations of stereophotographs at the time of imaging by 2 independent masked graders	SD-OCT RNFL imaging compared with confocal scanning laser ophthalmoscopy (CSLO)	RNFL assessment with SD-OCT performed well in detecting preperimetric glaucomatous damage in glaucoma suspects and performed better than CSLO, as measured by AUC, *however, the authors conclude that the test may not be the most appropriate for screening for glaucomatous visual field loss in the general population*
Pomorska 2011[15] Poland Prospective cohort study	27 eyes with ocular hypertension (OHT), 33 eyes with pre-perimetric glaucoma (PG), 30 perimetrically unaffected eyes of patients with glaucoma in the other eye (OE), and 58 eyes of age-matched healthy controls	Inclusion: Age≥18, Visual acuity 6/12 or better, spherical refractive error within ±5.0D and cylinder correction within ±3.0D, open angle on gonioscopy and a normal visual field Exclusion: Hx of intraocular surgery, ocular	Standard ophthalmic exam, including visual field testing	Time domain (TD)-OCT of optic nerve head (ONH) and RFNL	RNFL parameters discriminated OHT and OE groups better than did ONH parameters. ONH parameters demonstrated better dx precision in differentiating PG

Study **Region** **Study Design**	**Populations (n, mean age, gender number, %)**	**Inclusion criteria** **Exclusion criteria**	**Reference standard(s)** **Tests carried out and interpreted by**	**Device**	**Findings**
		trauma, diabetes, neurological disorders affecting visual field, pathological changes in the posterior segment, unreliable visual field results or poor quality OCT images			and normal eyes.
Kaushik 2011[11] India Prospective cohort study	123 subjects: 68 healthy, 32 glaucoma suspects, and 23 with glaucoma	Inclusion: POAG patients were included if Visual acuity 20/40 or better, spherical refractive error within ±5.0D and cylinder correction within ±3.0D; open angles on gonioscopy, characteristic glaucomatous optic neuropathy (GON) (cup/disc ratio >0.6), any diffuse or focal rim thinning, disc hemorrhage, and/or RFNL defects (with corresponding visual field defects defined as mean defect and pattern standard deviation outside 95% normal confidence limits and glaucoma hemifield test outside normal limits. Normals were required to have BCVA>20/40, IOP<21, open angles, normal-appearing optic disc, and normal visual fields. Eyes with optic discs suggestive of GON were considered glaucoma suspects. Exclusion: Hx of intraocular surgery, ocular disease, or eye with media clarity resulting in less than 20/40 or better view of the retina	Comprehensive ophthalmic exam Experienced OCT operator masked to patient's diagnosis	Cirrus® SD-OCT vs. Stratus® TD-OCT	AUC were comparable for both the SD and TD-OCT. RFNL thickness measures correlate well but do not have clinically acceptable agreement and therefore are not interchangeable. (not clear this study applies to screening)

Study **Region** **Study Design**	**Populations (n, mean age, gender number, %)**	**Inclusion criteria** **Exclusion criteria**	**Reference standard(s)** **Tests carried out and interpreted by**	**Device**	**Findings**
Chen 2012[4] Republic of China (Taiwan) Prospective cross-sectional study	161 participants: 21 OH patients, 27 glaucoma suspects, 35 patients with POAG, 26 patients with PCAG, and 52 healthy	Inclusion: Visual acuity 20/40 or better, spherical refractive error within ±5.0D and cylinder correction within ±3.0D; Glaucoma patients had to have GON (>0.2 cup/disc asymmetry between eyes, rim thinning, notching, excavation, or RNFL defect. and associated visual field loss. POAG patients had to have an initial IOP of >21mm (GAT) or the presence of open angle on gonioscopy, typical GON, and reproducible glaucomatous visual field defect with no other potential explanation. Exclusion: Co-existing retinal disease, uveitis, non-glaucomatous optic neuropathy	Complete ophthalmic exam	Cirrus® SD-OCT vs. Stratus® TD-OCT	Both SD- and TD-OCT showed comparable diagnostic power, e.g., to distinguish early glaucoma, OH, glaucoma suspects, POAG, and PACG in a Taiwanese population.
Bengtsson 2012[20] Sweden Prospective cohort study	307 randomly selected residents > 50 years of age living in 2 rural areas and 394 randomly selected glaucoma patients (open angle or pseudo exfoliation); of those, 138 healthy residents and 138 glaucoma patients (mean age 71 years) participated	For glaucoma patients, glaucomatous changes in optic disc, confirmed at study visit For population-based participants, those with non-glaucomatous disease known to affect the optic disc were excluded	Complete ophthalmic exam	SD-OCT vs. TD-OCT of RNFL thickness (average thickness, quadrant, and clock hour) for screening	Of 138 healthy residents, 9 were diagnosed with glaucoma. Stratus (TD) OCT had better specificity and PPV than Cirrus SD-OCT. Sensitivity, and therefore early detection, was better with Cirrus. Average RNFL thickness:

Study **Region** **Study Design**	**Populations (n, mean age, gender number, %)**	**Inclusion criteria** **Exclusion criteria**	**Reference standard(s)** **Tests carried out and interpreted by**	**Device**	**Findings**
					Stratus 100% specificity, 100% PPV, 68% sensitivity among whole group of glaucoma patients but only 28% sensitivity among earliest stage patients. Thus Stratus seems to be better choice for screening
OCT vs. other tests					
Pablo 2011[12] Argentina Cross sectional study	220 glaucoma suspects with ocular hypertension referred to a specialty clinic from outpatient clinics, of whom 181 (eyes)participated in study	Inclusion: Best corrected visual acuity ≥20/30; open anterior chamber angle on gonioscopy; spherical refractive error within ±5.0D and cylinder correction within ±2.0D (astigmatism); transparent ocular media; >18 years of age; IOP ≥22 on at least 3 separate visits, normal SAP (standard automated perimetry) in both eyes Exclusion: history of retinal pathology, retinal laser procedure; incisional ocular surgery, or neurological disease	Complete ophthalmic exam including visual field testing and RNFL monochromatic photography	TD-OCT to measure RNFL thickness GDx-VCC scanning laser polarimetry	OCT at average thickness had the highest diagnostic accuracy for detecting early RNFL defects (AUC 0.785) GDxVCC parameter was temporal-superior-nasal-inferior-temporal average (AUC 0.758); thus both procedures had similar dx accuracy for identifying early defects detected by RNFL photography in patients with OHT
Garas 2012[16]	177 consecutive	Inclusion:	Complete	Fourier Domain(FD)-	All 3 of the methods

Study **Region** **Study Design**	**Populations (n, mean age, gender number, %)**	**Inclusion criteria** **Exclusion criteria**	**Reference standard(s)** **Tests carried out and interpreted by**	**Device**	**Findings**
Hungary Prospective cohort study	individuals referred for glaucoma dx by family physicians: 50 healthy patients 28 OHT 33 preperimetric 66 perimetric	Healthy individuals: Refractive error within ±10.0D, sufficient central vision for optimal fixation, image quality sufficient for optimal evaluation, and no cataract Patients: no ONH damage, reliable and reproducible normal visual field with MD<2dB and untreated IOP>21mm	ophthalmic exam	OCT of RNFL thickness vs. GDx-VCC/ECC scanning laser polarimetry	were highly specific but for localized RNFLT damage, OCT was significantly more sensitive than GDx-VCC and GDx-ECC. 87.7% of cases were diagnosed by all 3 methods
Screening Approaches					
Campbell 2012[17] UK Delphi process	Using a Delphi process, 38 experts asked to nominate 3 screening interventions (a combination of 4 domains: target population, site, screening test, and test operator. 2-round process	Not relevant	Not relevant	Not relevant	Using a cutoff of a median rating of feasibility of ≥5.5, 6 interventions were identified. Beginning at age 50, using 2 or 3 screening tests (varying combinations of tonometry (IOP)/measures of visual function (e.g., FDT)/ optic nerve damage(e.g., HRT or OCT and cup/disc ratio)), organized in a community setting with an ophthalmic trained technical assistant as tester. Alternative was

Study Region Study Design	Populations (n, mean age, gender number, %)	Inclusion criteria Exclusion criteria	Reference standard(s) Tests carried out and interpreted by	Device	Findings
					glaucoma risk score based on questionnaire.

Table Notes: DDLS Disk Damage Likelihood Scale; FD Fourier Domain; FDT frequency doubling technology; GPS glaucoma probability score; HRT Heidelberg Retina Tomograph; MGSCP Mobile Glaucoma Screening Clinic Project; MRA Moorfields regression analysis; NFL nerve fiber layer; OAG open-angle glaucoma; OCT optical coherence tomography; OHT ocular hypertension; SD spectral domain; TD time domain

Appendix C. Questionnaire Matrix

Surveillance and Identification of Triggers for Updating Systematic Reviews for the EHC Program

Title: Screening for Glaucoma: Comparative Effectiveness

Conclusions From CER Executive Summary	Is this conclusion almost certainly still supported by the evidence?	Has there been new evidence that may change this conclusion?	Do Not Know
KQ1a: Does a screening-based program for open-angle glaucoma lead to less visual impairment when compared with no screening program? KQ1b: How does visual impairment vary when comparing different screening-based programs for open-angle glaucoma?			
We did not identify any study that addressed whether participation in an OAG screening-based program leads to less visual impairment when compared with no screening or another screening-based program.	☐	New Evidence:	☐
KQ2a: Does a screening-based program for open-angle glaucoma lead to improvements in patient-reported outcomes when compared to no screening? KQ2b: How do patient-reported outcomes vary when comparing different screening-based programs for openangle glaucoma?			
We did not identify any study that addressed whether participation in an OAG screening-based program leads to improvements in patient-reported outcomes when compared with no screening or another screening-based program.	☐	New Evidence:	☐
KQ 3: What is the predictive value of screening tests for openangle glaucoma?			
Evidence for this question came from two sources: a 2007 systematic review of diagnostic test accuracy for OAG (with subsequent diagnosis on followup as the reference standard) and original studies published subsequent to that review. No RCTs were identified. The sensitivity and	☐	New Evidence:	☐

Conclusions From CER Executive Summary	Is this conclusion almost certainly still supported by the evidence?	Has there been new evidence that may change this conclusion?	Do Not Know
specificity are reported by test, based on the evidence review and the subsequent literature. The authors of the original SR noted significant heterogeneity and high risk of bias for most groups of studies.			
Tests of Optic Nerve Structure **Heidelberg Retina Tomograph II** **Evidence From Burr et al., 2007**. pooled sensitivity: 86 percent (95% credible interval [CrI], 55 to 97) pooled specificity 89 percent (95% CrI, 66 to 98). **Evidence From Primary Studies.** Of 17 primary studies, 2 specifically focused on detecting early or moderate glaucoma. Sensitivity across 12 parameters: 47.1 percent (RNFL cross-sectional area) to 74.3 percent (linear cup/disc area ratio), Specificity: 47.1 percent (mean RNFL thickness) to 71.4 percent (cup shape measure). 15 studies compared HRT II with other devices. HRT II was found not to perform as well as GDx VCC, OCT, or FDT. HRT II and HRT III were found to have a similar diagnostic profile. Three of the included studies concluded that HRT II was not an appropriate tool	☐	New Evidence:	☐
Heidelberg Retina Tomograph III **Evidence from Burr et al.:** not included **Evidence From Primary Studies.** Eleven studies examined the diagnostic accuracy of HRT III. The sensitivity of the Glaucoma Probability Score (HRT III) for distinguishing eyes with early field loss from healthy eyes was 67.9 percent, and that of the Moorfields Regression Analysis was 71.9 (at a fixed specificity of 92 percent). The investigators concluded that "the relative ease and sensitivity of the operator-independent Glaucoma Probability Score function of the HRT III may facilitate glaucoma screening." In another study, sensitivity (parameter: reference height) ranged from 4 to 70 percent when holding the specificity of the test constant at 95 percent.	☐	New Evidence:	☐

Conclusions From CER Executive Summary	Is this conclusion almost certainly still supported by the evidence?	Has there been new evidence that may change this conclusion?	Do Not Know
Ophthalmoscopy **Evidence From Burr et al., 2007**. Seven studies addressed the diagnostic accuracy of ophthalmoscopy. Pooled sensitivity for five studies with common cutoff point of a vertical cup-to-disc ratio greater than or equal to 0.7: 60 percent (95% CrI, 34 to 82) Specificity: 94 percent (95% CrI, 76 to 99). The diagnostic odds ratio (DOR) was 25.7 (95% CrI, 5.79 to 109.50), suggesting a 26-fold higher odds of a positive test among those with glaucoma than those without glaucoma.	☐	New Evidence:	☐
Optical Coherence Tomography (OCT) **Evidence From Primary Studies**. 47 included studies that investigated the diagnostic accuracy of OCT: 34 considered the Stratus OCT, 10 included the Cirrus OCT, 6 considered the RTVue OCT, 2 included the Spectralis OCT, 2 examined the OTI OCT, and 1 included the OTI Spectral OCT/SLO. Stratus OCT studies at high risk of spectrum bias because those with known disease as well as those with healthy eyes were enrolled in the studies. Range of specificity: 66 to 100 percent.	☐	New Evidence:	☐
Optic Disc Photography **Evidence From Burr et al., 2007.** (6 studies) Sensitivity: 65 to 77 percent, Specificity: 59 to 98 percent. **Evidence From Primary Studies.** two studies of the diagnostic accuracy of optic disc photography and one study of cup-to-disc ratio measurement as measured by an ophthalmologist using a slit-lamp biomicroscope and 78 Diopter lens. Danesh-Meyer et al. (2006) included participants with OAG as well as glaucoma suspects and healthy volunteers. The AUC (comparison of those deemed to have glaucoma and borderline disease vs. normal) was 0.84 (95% confidence interval [CI], 0.74 to 0.92) for the cup-to-disc ratio and 0.95 (95% CI, 0.80 to 0.98) for the Disc Damage Likelihood Score, suggesting that the Disc	☐	New Evidence:	☐

Conclusions From CER Executive Summary	Is this conclusion almost certainly still supported by the evidence?	Has there been new evidence that may change this conclusion?	Do Not Know
Damage Likelihood Score is a more effective means of discriminating people with and without disease. The diagnostic accuracy of cup-to-disc ratio measurement from the Francis et al. (2011) study is described in the section on FDT C-20 perimetry.			
RNFL Photography **Evidence From Burr et al., 2007.** (4 studies) common cutoff point was diffuse and/or localized defect observed on RNFL photographs. Pooled diagnostic OR 23.1 (95% CI, 4.41 to 123.50) Pooled sensitivity: 75 percent Pooled specificity: 88 percent **Evidence From Primary Studies.** Two studies examined the accuracy of RNFL photography. Hong et al. (2007) analyzed RNFL photographs of 72 glaucoma and 48 healthy participants. Results showed the RNFL defect score II, with an AUC of 0.75 ($p < 0.001$), was the best parameter for discriminating early glaucoma from healthy eyes (sensitivity, 58.3 percent; specificity, 95.8 percent). Medeiros et al. (2004) compared RNFL photography with the GDx with VCC in 42 participants with OAG, 32 persons suspected of having OAG, and 40 healthy volunteers. The sensitivities of the global RNFL score were 36 and 81 percent, respectively, for fixed specificities of 95 and 80 percent. At a fixed specificity of 95 percent, the sensitivity of the Nerve Fiber Indicator was 71 percent versus the 36 percent reported above for red-free photos. Overall, the global RNFL score determined from red-free photos did not perform as well as scanning laser polarimetry. The AUC was 0.91 for the GDx with VCC Nerve Fiber Indicator versus 0.84 for the global RNFL score.	☐	New Evidence:	☐
Scanning Laser Polarimetry (GDx) **Evidence From Primary Studies.** Twenty-seven	☐	New Evidence:	☐

Conclusions From CER Executive Summary	Is this conclusion almost certainly still supported by the evidence?	Has there been new evidence that may change this conclusion?	Do Not Know
studies included an investigation of the GDx with VCC. The aim 12 of eight studies was to discriminate early glaucoma from no disease. In the studies that focused on early OAG, the range of sensitivity across all comparisons and cutoffs for the most frequently reported parameter—Temporal, Superior, Nasal, Inferior, Temporal average—was 29.8 to 81.63 percent. Specificity was fixed at 80, 90, or 95 percent in three studies, and the lowest reported specificity was 66.36 percent. The range in sensitivity for the nerve fiber indicator parameter across all comparisons and cutoffs was from 28.3 to 93.3 percent. The lowest specificity reported was 52.9 percent or was fixed at 80, 90, or 95 percent. Three studies examined the GDx with enhanced corneal compensation (ECC). The sample sizes of the included studies ranged from 63 to 92 glaucoma participants and 41 to 95 healthy volunteers. Medeiros et al. (2007) compared the AUCs for GDx with VCC and GDx with ECC, and reported that GDx with ECC performed significantly better than GDx with VCC for the parameters Temporal, Superior, Nasal, Inferior, Temporal average, Superior average, and Inferior average (p = <0.01). Sehi et al. (2007) and Mai et al. (2007) concurred with Medeiros et al. (2007) that imaging with ECC appears to improve the ability to diagnose OAG.			
Tests of Optic Nerve Function ***FDT (C-20-1) Perimetry*** **Evidence From Burr et al., 2007.** The pooled sensitivity and specificity results for the three studies that included FDT (C-20-1) perimetry and the common diagnostic criterion of one abnormal test point were high (92 and 94 percent, respectively). **Evidence From Primary Studies.** Four studies discussed the accuracy of FDT C-20 perimetry.	☐	New Evidence:	☐

Conclusions From CER Executive Summary	Is this conclusion almost certainly still supported by the evidence?	Has there been new evidence that may change this conclusion?	Do Not Know
Pueyo et al. (2009) enrolled 130 participants with ocular hypertension and 48 healthy volunteers. Using a cutoff of a cluster of at least four points with a sensitivity outside 95 percent normal limits, or three points outside 98 percent normal limits, or at least one point outside 99 percent normal limits, investigators determined the sensitivity of FDT to be 31.25 percent and its specificity 72.9 percent among the subset of 32 participants with glaucomatous optic neuropathy (of the 130 with ocular hypertension). The investigators concluded that FDT might not be an ideal test for participants with early defects. Salim et al. (2009) enrolled 35 participants with known OAG and 35 age- and sex-matched controls with no evidence of glaucoma. Investigators used FDT, noncontact tonometry, and a questionnaire individually and in all possible combinations to determine the accuracy of single and combination tests. FDT's sensitivity was 58.1 percent and its specificity was 98.6 percent. Overall, FDT was determined to be the best among the candidate single and combination tests in the study, despite fair sensitivity for detecting OAG. Pierre-Filho et al. (2006) enrolled glaucoma patients who had never experienced perimetry prior to the study. The investigators reported that 21 (32.8 percent) of the 64 participants with glaucoma were identified as having early disease, but data were not provided for this subgroup. Sensitivity and specificity were 85.9 and 73.6 percent, respectively, for the presence of at least one abnormal location and 82.8 and 83 percent, respectively, for two or more abnormal locations, regardless of severity. Francis et al. (2011) conducted population-based screening of 6,082 Latinos age 40 years and older as part of the Los Angeles Latino Eye Study			

Conclusions From CER Executive Summary	Is this conclusion almost certainly still supported by the evidence?	Has there been new evidence that may change this conclusion?	Do Not Know
(LALES) to determine the diagnostic accuracy of candidate screening tests performed alone or in combination. Participants completed Humphrey Visual Field testing in addition to FDT C-20-1, GAT, and central corneal thickness and cup-to-disc ratio measurements. Diagnostic test accuracy outcomes were assessed for the general population as well as high-risk subgroups, defined as persons who were 65 years and older, those with a family history of glaucoma, and persons with diabetes. Of the 6,082 participants screened, 4.7 percent (286) were diagnosed as having OAG. Based on three glaucoma diagnosis definitions (glaucomatous optic nerve appearance, glaucomatous visual field loss, glaucomatous optic nerve and visual field loss), the test parameters vertical cup-to-disc ratio ≥ 0.8 and Humphrey Visual Field (HVF) false negatives ≥ 33 percent had the highest specificity, regardless of the definition of glaucoma (98 percent). HVF mean deviation < 5 percent had the highest sensitivity (78 percent) using the definition of optic nerve defects only, while the HVF glaucoma hemifield test had the highest sensitivity under the other two definitions (90 percent for glaucomatous visual field loss and 90 percent for both field loss and optic nerve damage). Specific results for the FDT C-20-1 were as follows (sensitivity/specificity, definition of glaucoma): 59 percent/ 79 percent, glaucomatous optic nerve appearance only; 68 percent/80 percent, glaucomatous visual field loss only; 67 percent/79 percent, both glaucomatous optic nerve appearance and visual field loss. The investigators reported similar results when high-risk subgroups were analyzed and concluded that "these results suggest that screening of high-risk groups based on these criteria may not improve over screening of the general population over age 40."			
FDT (C-20-5) Perimetry **Evidence From Burr et al., 2007.** Five studies of FDT (C-20-5) with significant heterogeneity using	☐	New Evidence:	☐

Conclusions From CER Executive Summary	Is this conclusion almost certainly still supported by the evidence?	Has there been new evidence that may change this conclusion?	Do Not Know
the common cutoff point of one abnormal test point were included. The range of sensitivity was 7 to 100 percent; the specificity range was 55 to 89 percent. ***FDT 24-2 Perimetry*** **Evidence From Primary Studies.** Five studies examined the diagnostic accuracy of FDT 24-2 threshold tests using the Humphrey Matrix Perimeter. All studies included participants with known glaucoma and healthy volunteers, and we judged these studies to be at high risk of spectrum bias. The range of sample size was 25 to 174 glaucomatous eyes and 15 to 164 healthy eyes. Sensitivities and specificities were reported for the parameters mean deviation, pattern standard deviation, and glaucoma hemifield test outside of normal limits. There was appreciable heterogeneity in the estimates of sensitivity at 80 percent, 90 percent, and 95 percent specificity that may be attributed to a number of factors, including different patient populations and variations in cutoff points. The sensitivity was 55 percent for the mean deviation and 94 percent at 80 percent fixed specificity. Tafreshi et al. (2009) and Leeprechanon et al. (2007) reported 39 and 87 percent at 90 percent fixed specificity, and 32 and 82 percent at fixed 95 percent specificity, respectively. Sensitivity and specificity for pattern standard of deviation (PSD) and glaucoma hemifield test are reported with their cutoff points in the evidence tables in Appendix C of the full report. Bagga et al. (2006) and Burgansky-Eliash et al. (2007) reported the AUC for the mean deviation parameter (0.69 for both studies with $p < 0.04$ and 95% CI, 0.564 to 0.815, respectively). The AUCs for PSD were 0.66 ($p = 0.09$) and 0.733 (95% CI, 0.618 to 0.848).			

Conclusions From CER Executive Summary	Is this conclusion almost certainly still supported by the evidence?	Has there been new evidence that may change this conclusion?	Do Not Know
FDT 30-2 Perimetry **Evidence From Primary Studies.** Two studies discussed the detection of early glaucoma using the FDT 30-2 threshold test with the Humphrey Matrix Perimeter. Both Hong, Chung, Hong, et al. (2007) and Hong, Ahn, Ha, et al. (2007) enrolled OAG participants with early visual field loss and healthy controls. The mean deviation and PSD were judged to be good parameters for distinguishing between eyes with early disease and eyes with no known defects. The mean deviations were 0.795 and 0.750 and the PSDs were 0.808 and 0.934 for Hong, Chung, Hong, et al. and Hong, Ahn, Ha, et al., respectively. Both study groups, however, determined that the best parameter for distinguishing eyes with early glaucoma from healthy eyes was the number of points that have p less than 5 percent in the pattern deviation plot, with an AUC of 0.985 (95% CI, 0.943 to 0.998) in Hong, Chung, Hong, et al. and 0.990 ($p < 0.001$) in Hong, Ahn, Ha, et al.	☐	New Evidence:	☐
FDT N-30 Perimetry **Evidence From Primary Studies.** Four studies examined the accuracy of the FDT N-30 threshold test. Zeppieri et al. (2010) focused on the detection of early glaucoma among a sample of 75 participants with OAG, 87 with ocular hypertension, 67 with glaucomatous optic neuropathy, and 90 healthy volunteers. At the best cutoff of less than -0.78, the sensitivity of the mean deviation parameter was 61.3 percent and the specificity was 73.7 percent for distinguishing early OAG from healthy eyes. At the best cutoff of greater than 3.89, the sensitivity of the PSD was 76.0 percent and the specificity was 87.8 percent. Salvetat et al. (2010) focused on the detection of early disease among a sample of 52 participants with early OAG and 53 healthy volunteers. The sensitivity of mean deviation for distinguishing early OAG from healthy eyes at the best cutoff (less than -1.12) was 67 percent and the specificity	☐	New Evidence:	☐

Conclusions From CER Executive Summary	Is this conclusion almost certainly still supported by the evidence?	Has there been new evidence that may change this conclusion?	Do Not Know
was 74 percent. At the best cutoff of greater than 3.97, the sensitivity of the parameter PSD was 96 percent and the specificity was 85 percent.			
Goldmann Applanation Tonometry **Evidence From Burr et al., 2007.** At the common cutoff point of IOP greater than 20.5-22 mm Hg, nine studies with significant heterogeneity reported sensitivity in the range of 10 to 90 percent and specificity in the range of 81 to 99 percent. **Evidence From Primary Studies.** Two studies included examination of GAT. Bagga et al. (2006) compared the ability of various tests of structure and function to discriminate healthy eyes (n = 22) from eyes with known glaucomatous optic neuropathy (n = 25). The AUC for IOP, as measured by GAT, was 0.66 (p = 0.05). The methods of the Francis et al. (2011) study (LALES) are discussed in the FDT C-20 section of the full report. The specific sensitivity and specificity values for GAT using a cutoff of ≥ 21 mm Hg for the three definitions of glaucoma were as follows (sensitivity/specificity, definition of glaucoma): 21 percent/97 percent, glaucomatous optic nerve appearance only; 23 percent/97 percent, glaucomatous visual field loss only; 24 percent/97 percent, both glaucomatous optic nerve appearance and visual field loss.	☐	New Evidence:	☐
Humphrey Visual Field Analyzer **Evidence From Primary Studies.** Ten studies examined the diagnostic accuracy of the HFA. Of these, six examined HFA Short Wavelength Automated Perimetry; two tested HFA-SAP, SAP-SITA, and HFA SAP-Full Threshold (FT); four examined HFA-SITA-Standard; and one tested the HFA SITA-Fast protocol. The HFA Short Wavelength Automated Perimetry testing protocol (the most frequently reported) included 25 to 286 participants with glaucoma and 22 to 289 healthy volunteers across the six included studies.	☐	New Evidence:	☐

Conclusions From CER Executive Summary	Is this conclusion almost certainly still supported by the evidence?	Has there been new evidence that may change this conclusion?	Do Not Know
Sensitivity across all comparisons and cutoffs for the mean deviation ranged from 25.9 to 83 percent. Specificity ranged from 80 to 95.2 percent. Cutoff points ranged from -5.42 to -11.06 dB.			
Noncontact Tonometry **Evidence From Burr et al., 2007.** (1 study) Sensitivity: 92 percent Specificity 92 percent (using the criterion of IOP greater than 21 mm Hg). **Evidence From Primary Studies.** Salim et al. (2009) included noncontact tonometry, individually and in all possible combinations, with other measures of structure and function to determine the accuracy of single and combination tests. IOP, as measured by noncontact tonometry, was found not to be a very sensitive test for detecting glaucoma (sensitivity 22.1 percent). The investigators acknowledge that use of topical medications by the glaucoma participants could limit the ability to identify those with disease.	☐	New Evidence:	☐
Oculokinetic Perimetry **Evidence From Burr et al., 2007.** Four studies were included that examined the diagnostic accuracy of oculokinetic perimetry. The common criterion varied in description, but is best described as one or more points missing. The odds of a positive test were 57 times as high (DOR, 57.54) for those with glaucoma as for those without glaucoma (95% CrI, 4.42 to 1585.00). The pooled sensitivity and specificity were 86 and 90 percent, respectively.	☐	New Evidence:	☐
SAP Suprathreshold Test **Evidence From Burr et al., 2007.** Nine studies, including the Baltimore Eye Survey and the Blue Mountains Eye Study, examined the SAP suprathreshold test. Although the sensitivity and specificity were similar for the Baltimore and Blue Mountains studies, there was significant heterogeneity among the included studies. The range in sensitivity was 25 to 90 percent; the range in specificity was 67 to 96 percent.	☐	New Evidence:	☐

Conclusions From CER Executive Summary	Is this conclusion almost certainly still supported by the evidence?	Has there been new evidence that may change this conclusion?	Do Not Know
SAP Threshold Test **Evidence From Burr et al., 2007.** Among the five studies analyzed for SAP threshold, both Humphrey 30-2 and 24-2 threshold and Octopus 500 were evaluated. The pooled sensitivity was 88 percent, and specificity was 80 percent for the common cutoff point. (The definition of the common cutoff point differed by included study, but is defined in Burr et al.)	☐	New Evidence:	☐
Tendency-Oriented Perimetry **Evidence From Primary Studies.** Pierre-Filho et al. (2006) compared frequency doubling technology), tendency-oriented perimetry using the Octopus 301 G1-TOP program, SITA Standard, and SITA Fast in 117 eyes (64 with glaucoma and 53 healthy eyes). The Octopus 301 perimeter test was considered abnormal under two conditions: when the mean defect was "> 2dB and/or the loss variance > 6 dB (TOP 1), and…there were at least seven points (three of them contiguous) with a reduction in sensitivity ≥ 5 dB in the corrected comparisons graphic (TOP 2)."92 The sensitivity using definition TOP 1 was 87.5 percent (95% CI, 76.3 to 94.1) and the specificity was 56.6 percent (95% CI, 42.4 to 69.9). With definition TOP 2, the sensitivity was 89.1 percent (95% CI, 78.2 to 95.1) and the specificity was 62.3 percent (95% CI, 47.9 to 74.9).	☐	New Evidence:	☐
Direct Comparisons of Candidate Tests **Evidence From Burr et al., 2007** Six studies included comparisons of SAP with optic disc photography, HRT II, FDT, and/or GAT. Burr et al. concluded that sensitivity results at the common cutoff point for each test revealed that SAP performed better than GAT. One of the two studies that addressed the comparison of SAP to GAT reported estimates of sensitivity of 89 percent and 3 to 14 percent, respectively. Specificity values were 73 percent for SAP and 98 to 99 percent for GAT. Burr et al. also concluded that SAP was similar to HRT II. The sensitivity of	☐	New Evidence:	☐

Conclusions From CER Executive Summary	Is this conclusion almost certainly still supported by the evidence?	Has there been new evidence that may change this conclusion?	Do Not Know
SAP was 72 percent and the sensitivity of HRT II was 69 percent in one of the two included studies; the specificity for both tests was 95 percent. There was one included study in which the investigators compared SAP with optic disc photography. Optic disc photographs had a similar sensitivity (73 to 77 percent) and specificity (59 to 62 percent) to SAP (sensitivity, 50 to 71 percent; specificity, 58 to 83 percent). In the two studies that included comparisons of SAP with FDT, one study reported similar sensitivity estimates (SAP, 63 to 90 percent; FDT C-20-5, 68 to 84 percent) and similar specificity values (SAP, 58 to 74 percent; FDT C-20-5, 55 to 76 percent). Based on analyses of the common criterion for each test, test accuracy, combination tests, tests for glaucoma at specific stages, and direct and indirect comparisons of tests, Burr et al. (2007) concluded that optic disc photography, HRT II, FDT, SAP, and GAT were candidates for use in a screening-based program.			
Conclusion Based on the Burr et al. (2007) findings, standard automated perimetry was compared with other tests available at that time. SAP had higher sensitivity than Goldmann tonometry, similar sensitivity to HRT, and lower sensitivity than disc photos or FDT. In terms of specificity, SAP performed better than disc photos and FDT, similar to HRT, and worse than Goldmann tonometry. **Evidence from Primary Studies:** Several additional studies assessed the performance of glaucoma screening tests not included in the Burr et al. review. The studies included newer imaging (GDx, HRT III, OCT) and functional (Short Wavelength Automated Perimetry, new FDT patterns) technologies. However, despite improvements in the technology, it is still not clear that there is any one test or combination of tests suitable for use in glaucoma screening in the general population.	☐	New Evidence:	☐

Conclusions From CER Executive Summary	Is this conclusion almost certainly still supported by the evidence?	Has there been new evidence that may change this conclusion?	Do Not Know
Significant barriers to identifying and characterizing potential glaucoma screening tests remain. These barriers include the lack of a definitive diagnostic reference standard for glaucoma and heterogeneity in the design and conduct of the studies. Because of these barriers, the ranges of sensitivities, specificities, and AUCs are large and prevent a coherent synthesis.			
KQ4a: Does a screening-based program for open-angle glaucoma lead to reductions in intraocular pressure when compared with no screening program? **KQ4b: How does intraocular pressure vary when comparing different screening-based programs for open-angle glaucoma?**			
No studies were identified that addressed whether participation in an OAG screening-based program leads to reductions in IOP when compared with no screening or another screening-based program.	☐	New Evidence:	☐
KQ5a: Does a screening-based program lead to a slowing of the progression of optic nerve damage and visual field loss when compared with no screening program? **KQ5b: How do optic nerve damage and visual field loss vary when comparing different screening-based programs for open-angle glaucoma?**			
Evidence From Systematic Reviews A systematic review of randomized trials of screening modalities for OAG compared with no screening (including opportunistic case finding and referral) (Hatt et al., 2006). There were no restrictions on included populations. The primary outcome of interest was the prevalence of visual field loss, defined as the proportion of participants with a prespecified severity of visual field loss diagnosed by either manual or automated field assessment. Other primary outcomes included the prevalence of optic nerve damage and visual impairment. Electronic searches of five databases, including MEDLINE and CENTRAL, were conducted in 2006 and again in January 2009, but none of the studies that were identified were eligible for inclusion. The review authors acknowledged that RCTs require lengthy followup and are predicated on identifying appropriate candidate tests that may be incorporated into a	☐	New Evidence:	☐

Conclusions From CER Executive Summary	Is this conclusion almost certainly still supported by the evidence?	Has there been new evidence that may change this conclusion?	Do Not Know
screening-based program. **Detailed Analysis of Primary Studies** We did not identify any study that addressed whether participation in an OAG screening-based program leads to reductions in visual field loss or optic nerve damage when compared with no screening or another screening-based program.			
KQ 6: What are the harms associated with screening for openangle glaucoma?			
We did not identify any study addressing the harms associated with screening for OAG.			
Are there new data that could inform the key questions that might not be addressed in the conclusions?			